LKMS/Mnemosyne: Semantic Web technologies and Markup techniques for a Legal Knowledge Management System.

L. Gilardoni, C. Biasuzzi, M. Ferraro, R. Fonti, P. Slavazza
Quinary
with F. Toffoletto, V. F. Giglio,
Toffoletto e Soci Studio Legale
The issue: handling the knowledge cycle

What a lawyer does?

- ... (re)build documents ...
  - based on knowledge

It is not just in the docs, it is around!

- Documents in Context
  - Why and When
    - Explicit (within documents)
    - Implicit
      - From the process
      - From Background knowledge

- People in Context
  - Who and Why.
    - From Workgroup
    - From Social Networks
    - To Knowledge Networks

- Going inside the process
  - Enabling building a culture of Knowledge Sharing,
  - to leverage a firm collective wisdom

© 2006 Quinary SpA
Key Issues

• Consolidating collective wisdom!
  • Documents + Context
    • Sharing Repository - Documents – in the wider sense
    • Semantic Markup - By Document Content - By Process Context
    • Intelligent Drafting - Codified Corporate Knowledge
  • Ease Sharing
    • Ease of usage; adding as a snap
    • Tailored Interfaces - Exploiting Background Knowledge
  • Collecting Context
    • From document content. Markup By Automatic Information Extraction
    • From process – who did what
    • Internal AND External Knowledge

• Semantic Web & Markup Technologies as a core toolset
  • RDFs ontologies as a glue
  • Web Services architecture
  • XML based assembly & interface tailoring
LKMS/Mnemosyne - Baseline

• LKMS: Legal Knowledge Management System / Mnemosyne (Μνημοσύνη the personification of Memory)

• A web based repository where lawyers can SHARE and retrieve

• “Legal oriented”: tools supporting specific legal processes

• Jointly developed by Quinary and Toffoletto&Soci since 2002
  • Partly supported by EC Sponsored project DotKom
  • Ongoing R&D partly within EC project X-Media
  • Best Industrial Paper at ISWC 2005

© 2006 Quinary SpA
LKMS/Mnemosyne - features

- Advanced sharing environment
- ‘Conventional’ stuff
  - Nodes taxonomies (used to organize material & process)
  - Docs + Urls + Notes management – upload/download/search
  - People activity tracking
- With advanced features
  - Legal Process Support - Integrated Document Assembly and case management (shown later)
    - Legal Docs automatic drafting: from conditions to structured documents
    - XML drafting – based on Exari technology Consolidation of law firm knowledge
    - Integrated process – seamless context gathering
  - Rich ontology base metadata (shown later)
    - RDF(s) annotations repository
    - Semantic annotation to leverage structural context
    - Automatic Information Extraction
  - Advanced sharing features
    - Who did what / tracking / also relevant … - based on interactions
    - Linking Inside and Outside Knowledge – News watching, spidering, KZilla
- Based on an open architecture
  - Java based, portable, web services
  - Tailorable – ontology driven - web interface
- Demo / Details available for interested people
- More on advanced sharing features in a forthcoming paper (Ekaw 06)
Ontology Based Annotations

• SemantiK
  • a “generic” platform for high-level access to an RDF knowledge base,
  • a knowledge integration layer, supporting user or agent annotations
  • a web service architecture supporting extraction agents
  • a web GUI for maintenance of RDF annotations – tailored to end users (Lawyers)

• A legal Ontology – Legal SemantiK
  • **Laws**: most of the subdivision of the Italian legislation:
    • law’s (different kinds) articles, articles’ subparts (refs to Nir ‘Articolato’ – simplified)
  • **Legal Documents**:
    • Contracts, Legal Opinions and Sentences ,
    • plus supporting entities, like tribunals, judges, Italian high court decisions, etc.
  • **Juridical Concepts**:
    • ‘legal keywords’ as normally used by lawyers, relevant matters.
  • **Focus**: not on structuring docs but on enriching with annotations; the ‘final user’ (Lawyer) perspective
    • Building annotations (manually – GUI issues – AND automatically)
    • Integration – partial matches, normalisation

• A legal references extractor
SemantiK - supporting rich annotations

Ontology Driven Fuzzy Match

Wrappers & Brokers for IE tools
SemantiK – better search

Semantic(K) Links

Full Text + Semantic Search

Semantic(K) Queries / Services
Search and inference operations over the annotations are performed against **Sesame RDF database**.
A Knowledge Handler is in charge of handling actions regarding specific ontology classes (i.e. for matching specific class attributes)
A look inside SemantiK - architecture

The **Presentation Layer** is in charge of showing the annotations to the user (viewing and editing tasks). Rendering is done with XSL + CSS technology.
The **Knowledge Integration** component matches user inputs with existing resources, detects conflicts between the KB and to-be-inserted annotations, drives manual annotation wizards.

© 2006 Quinary SpA
A **Knowledge Broker** is an interface to an external IE component, in charge of extracting annotations from the documents automatically.
Legal SemantiK – Knowledge Handlers

- User input is dispatched to the right Knowledge Handler, depending on the range of the property in context, such as to find and match legal documents by number and date, as opposed to judges, matched primarily by name and surname.

- The Knowledge Handler for a law is also able to compute URNs, standard unique identifiers used to build hyperlinks to external resources, such as the Norme In Rete site.
Legal SemantiK– Complexity hiding

Semantic annotation wizards: exploiting ontology constraints.
Legal SemantiK – Annotation Extraction

- Automatic tagging by an Information Extraction application
  - Legal References (laws, sentences) from any document.
  - Document specific metadata for Sentences
- References highlighted on text
  - Offset to the document as XPointers
- Cross referencing & search through RDF
- ‘Standard Module’ based on GATE/JAPE
- Experimental ones with Adaptive Methods
  - R&D From DotKom, XMedia
- IE application functionalities are exposed as Web-Services
- Each IE engine is mediated in SemantiK by a Knowledge Broker

© 2006 Quinary SpA
SemantiK – Raw numbers

- Semantik extension to LKMS has been deployed during fall 2004 at our first customer (50 lawyers). Since then part of the core system
  - The system hosts currently about 50,000 documents.
  - All new documents and part of the old ones are being manually annotated with semantik
  - Automatic IE is going in production since last spring
- Deployed in 2005 at another Italian major law firm (200 users)
  - Automatic IE currently under evaluation
- Sentence archival services at AGI – since April 1°, all material going through automatic extraction (about 95% accuracy)
- Same architecture underlying non legal environments:
  - Quinary (our own intranet)
  - A major Italian utility company
  - EC Project X-Media
Sample from AGI service
SemantiK – Evaluation & Directions

• Qualitative evaluation from law firms with very good judgments and strong support from law firm management
• Formal Quantitative evaluation of automatic IE results performed within DotKom EC Project
  • IE extraction of legal references on a corpus of 1200 documents,
    • 5800 references to a total of 250 laws, 1150 articles to laws.
    • average precision > .85, recall ranging between .70 and .80.
• Recent experiments on about 800 court sentences (informal measures)
  • About 80%-90% accuracy at start
  • About 95% after some extension
• Ongoing R&D
  • Extending coverage (International law) & performance
  • Better management of ontology lifecycle, very large scale and multiple media (within EC sponsored X-Media)
The drafting Component

• A contract (or any document) as a XML structured clause set

• Logic Components
  • Conditions
  • Smart Phrases
  • Variables
  • Queries/Calculation

• A specialised XML editor
  • Full fledged XML editor
  • Binding text to logic
  • Style based L&F

• Able to import from Word
  • Also importing styles
About

• **Quinary**
  • Since 1985, named after: P. Hatt, N Foote "American Economic Review" May ‘53:
    • *quaternary* - advanced services in the financial sector
    • *quinary* - everything 'advanced' in the technology and R&D areas
    • towards an economy of professional services and knowledge exchanges
  • A long history providing advanced ICT solutions built around Knowledge Technologies
  • Since 2000 R&D efforts towards Knowledge and Law
  • [http://www.quinary.com](http://www.quinary.com)

• **Toffoletto e Soci**
  • A leading Italian Law firm, since 1927 specializing in Employment Law. Member of the IusLaboris international alliance.
  • [http://www.toffoletto.it](http://www.toffoletto.it), [http://www.iuslaboris.com](http://www.iuslaboris.com)

• **LKMS/Mnemosyne**
  • a Knowledge Management System tailored for the legal profession
  • [http://mnemosyne.quinary.com](http://mnemosyne.quinary.com)
References

- Parigi P., Martello S. “Organizzazione e Marketing dello Studio Legale” I libri di Guida Al Diritto, Il Sole 24 Ore 2006
- Slavazza P., Fonti R., Ferraro M., Biasuzzi C., Gilardoni L. “Towards a Knowledge Ecosystem”, To appera in the proceedings of Ekaw-06, 15th International Conference on Knowledge Engineering and Knowledge Management, Prague Oct. 2006